REVISITING THE KULESHOV EFFECT WITH FIRST-TIME VIEWERS

Sermin Ildirar[1, 2], Louise Ewing[1, 3]

[1] Department of Psychological Sciences, Birkbeck, University of London, UK

[2] Cinema Department, Istanbul University, Turkey

[3] School of Psychology, University of East Anglia, Norwich, UK

Dr. Sermin Ildirar is a Marie S. Curie post-doctoral research fellow at the Centre for Brain and Cognitive Development, Birkbeck, University of London. She studied Film and Media studies at Istanbul University and University of Vienna. She worked at the Knowledge Media Research Center in Tubingen, Germany. She is currently continuing to investigate the role of cinematic features on cognitive processes of the both adult and infant viewers. She has articles published in journals such as Psychological Science and British Journal of Psychology. She is also the director of several short movies and co-scriptwriter of a feature movie and a computer game.

Address: Centre for Brain & Cognitive Development Department of Psychological Sciences Birkbeck, University of London 32 Torrington Square London WC1E 7HX, U.K.

Tel: +44 (0)20 7079 0757 **Fax:** +44 (0)20 7631 6587 **Email:** ildirarsermin@gmail.com

Dr Louise Ewing is a lecturer at the University of East Anglia, Norwich. She studied Psychology at the University of Western Australia. Her research specialism is face and person perception. She has articles published in journals such as Perception, Vision Research and the Journal of Experimental Psychology.

Address: School of Psychology, University of East Anglia, Norwich Research Park, NR4 7TJ,, U.K. Tel: +44 (0) 1603 59 7379 **Email:** l.ewing@uea.ac.uk

Word count, total= 5947

Number of tables= 3

Number of Figures=2

Abstract

Researchers have recently suggested that historically mixed findings in studies of the Kuleshov effect (a classic film editing-related phenomenon whereby meaning is extracted from the interaction of sequential camera shots) might reflect differences in the relative sophistication of early vs. modern cinema audiences. Relative to experienced audiences, first-time film viewers might be less predisposed and/or able to forge the required conceptual and perceptual links between the edited shots in order to demonstrate the effect. The current study recreates the conditions that traditionally elicit this effect (whereby a neutral face comes to be perceived as expressive after it is juxtaposed with independent images: a bowl of soup, a gravestone, a child playing) to directly investigate and compare "continuity" *perception* in first-time and more experienced film viewers. Results confirm the presence of the Kuleshov effect for experienced viewers (explicitly only in the sadness condition) but not the first-time viewers, who failed to perceive continuity between the shots.

Key words: Kuleshov effect, continuity perception, artificial landscape, first-time viewers, naive viewers

Revisiting the Kuleshov Effect With First-Time Viewers

The Kuleshov effect is a film editing effect demonstrated during the late 1910s and early 1920s by the Russian pioneering filmmaker and theorist Lev Kuleshov (1899–1970). Famously, Kuleshov is reported to have intercut a close-up of the Russian actor Mozhukin's neutral, expressionless face with various other camera shots, including a bowl of soup, a woman in a coffin and a child playing with a toy bear. He observed that these additional shots interacted with the original, leading viewers to perceive the (objectively neutral) face as expressing happiness, sadness, and hunger/thoughtfulness respectively (Pudovkin 2007). As the years have passed, the reliability and validity of this effect have come into question. The original footage used by Kuleshov is long-since lost and superficial issues with the design of the experiments[1] have prompted some to re-classify it as part of the "mythology of film" (Holland 1989) or "folklore of the cinema" (Pearson and Simpson 2005). Yet this disapproval may be unwarranted.

Despite the somewhat anecdotal nature of Kuleshov's original observations, other (more rigorous) studies provide converging evidence that a single film scene can generate a profoundly different perceptual meaning for viewers when placed in different contexts. Goldberg (1951) for example, found that the emotional quality and intensity of a fearful face accompanied by a scream can differ depending on the order of camera shots, e.g., it can come to be perceived as rage or even joy. Similarly, studies by Kuiper (1958) and Foley (1966)

[1] It has been both (conflictingly) reported that Kuleshov found a long strip of film with Mozhukhin's close-up and used it for his experiment (Levaco, 1974, p.8); and that he purpose filmed Mozhukhin, with the instruction to appear expressionless (Messaris, 1994, p.16).

demonstrated that neutral faces can be perceived as happy or sad, depending on their context in films (as cited in Isenhour 1975). Support has also come from psychological studies utilizing brain imaging (Mobbs et al. 2006) and eye tracking (Aviezer et al., 2008; Barratt, Rédei, Innes-Ker, and van de Weijer, 2016) techniques during the viewing of edited film clips. Mobbs and colleagues (2006) observed differential neural responses (e.g., in the bilateral temporal pole, superior temporal sulcus and anterior cingulate cortices) when identical faces were paired with different emotionally salient contextual movies. At the end of the scanning session, they also asked their subjects to judge the faces. Despite faces being identical, attributions of facial expression and mental-state were altered when they were juxtaposed with contextual movies of different valance. Aviezer and collegues (2008) reported that the pattern of participants' eye movements to facial regions changed systematically as a function of the affective context in which these images appeared.

The most recent replication (and extension) of the Kuleshov experiment was conducted by Barratt and colleagues (2016) and concluded that "some sort of Kuleshov effect does in fact exist" (p.865). These authors considered the original film sequences to be an instance of point-of-view editing, so carefully constructed their set of test stimuli to encourage participants to infer that the glance shot and the object shot were spatially related, i.e., the gazer did not look directly into the camera. Their results confirmed that the emotional context influenced participants' judgments of the target face stimulus in each of the five emotional conditions (happiness, sadness, hunger, fear, and desire), with the most pronounced effects for sadness.

Importantly, however, previous replication attempts have been less successful. Prince and Hensley (1992) found that the majority of their subjects reported seeing an actor with a neutral expression (i.e., no editing-induced appearance of emotion), regardless of the

sequence into which his face was edited. These authors suggested that the "naiveté of early cinema audiences", compared with their more experienced, modern participants (university undergraduates), might explain the original findings.

To our knowledge there has been no empirical study of the Kuleshov effect with naive participants. However, there have been anecdotal reports (Forsdale & Forsdale, 1970) and direct investigations of their perception of other aspects of editing (e.g., Hobbs et al., 1988; Schwan & Ildirar, 2010). Hobbs and colleagues (1988) compared single shot recordings with edited versions of the same content and reported no effect of editing on comprehension in first time viewers. Crucially however, more recent studies with first-time viewers (Schwan & Ildirar ,2010; Ildirar & Schwan, 2014; Ildirar, Levin, Schwan and Smith, 2014) have found that participants' familiarity with the depicted content can powerfully modulate this effect. In their studies, first-time viewers struggled to construct a spatiotemporal relationship between adjacent shots (e.g., shot reverse shot, outdoor to indoor shot). Instead, they perceive adjacent camera shots as independent images unless there is an ongoing line of actions viewers are familiar with, a salient gaze cue or clear dialogue. Given that the film clips historically used in Kuleshov experiments do not include any such cues (relying instead on participants connecting the shots together through emotion), it remains an open question whether this editing effect will help naive viewers to perceive a spatiotemporal relationship between the adjacent shots. In order to answer this question, we conducted a field experiment that attempted to elicit the Kuleshov effect with a unique sample of first-time film viewers from regional Turkey, and a comparison group from the same region.

(Figure 1 around here)

Kuleshov-Type Sequence as an Instance of Artificial Landscape

There are two components to the Kuleshov effect: perception of spatiotemporal continuity between the juxtaposed camera shots and perception of a change in emotion of the target (neutral) face. Although the first component is a critical prerequisite for the latter, it is rarely directly considered or discussed in any detail. An exception is David Bordwell, Kristin Thompson and Jeremy Ashton's consideration: They call the Kuleshov effect any series of shots that in the absence of an establishing shot prompts the spectator to infer a spatial whole on the basis of seeing only portions of the space. Here the authors describe artificial landscape without naming it.

While shooting his film *The Project of Engineer Prite* (1918), Kuleshov discovered that it was possible to create a cinematic terrain that exists nowhere in reality. This was the first of several properties of montage he described in his later articles and books. His film required shots of actors looking at electrical cables strung on poles that had not been filmed. Kuleshov supposed that the same effect could be achieved by splicing shots of actors looking off-camera with separately recorded shots of the row of poles. Since the poles and the actors were in different parts of Moscow, Kuleshov termed the effect the "artificial landscape" (also known as "creative geography") (Kuleshov 1974). After this discovery, Kuleshov, created other *artificial landscapes* in his movies. For example, he presented scenes in which actors walked up the steps of a well-known Moscow building to then arrive at The White House in the USA. In one film he even combined close up shots of different women's body parts to create a "new" woman. In this way, he created cities, buildings and bodies that existed only on screen.

Artificial landscape is a ubiquitous feature of modern film and television. For example, when two characters are shown in single shots looking right and left respectively (usually in dialogue scenes) viewers readily assume that they are filmed in the same place at

the same time, though this may not have been the case. A well-known example is dialogue between David Bowie and Marlene Dietrich in *Just a Gigolo* (1978), which was filmed with these actors individually, in separate rooms, months apart. It is interesting to note that although the viewers of *Just a Gigolo* did not realize this production trick and perceive the shots in a spatiotemporal continuity, first-time film viewers were not similarly fooled (Schwan & Ildirar, 2010; Ildirar & Schwan, 2014). These naïve viewers saw people in the same place but not at the same time, i.e. thinking that the second person comes after the first one leaves.

Kuleshov-Type Sequence as an Instance of Point-of-View (POV) Shot

Another master of editing, Alfred Hitchcock, noted that the primary editing structure of his film Rear Window (1954) was based on the Kuleshov effect. In the film, James Stewart's character (Jeff) is a voyeur, peeking through his window into people's private lives. In the framing of the shots, Hitchcock consistently kept his POV shot aligned with Stewart's eyeline. Since Stewart often has an emotionally ambiguous face during the film, the views out of his apartment window powerfully drive the emotional context (Truffaut, 1984, pp. 213–223). In an interview, Stewart later claimed not to remember playing the role the way he had seen it on-screen. Thus it appears that Hitchcock's manipulation of the Kuleshov effect was so successful, he was able to alter the montage to create completely different meanings (Scharff, 1997).

From this perspective, a Kuleshov-type sequence can be considered as an instance of point-of-view (POV) shot, which is a short film scene that shows what a character (the subject) is looking at (represented through the camera). Viewers link these two images together in their minds and perceive them as they are depicting a continuous instance - concluding that the person is looking at the object.

The POV shot is one of the techniques film makers discovered in the early years of cinema, which helps viewers to integrate diverse views separated by cuts, in other words to perceive continuity through film cuts. One proposed explanation of how viewers perceive cinematic continuity in spite of the spatio-temporally discontinuous nature of the visual information presented to them is that the films produce a stream of audiovisual information that is similar to our veridical perception of real scenes and events (e.g., Munsterberg, 1916/1970; Anderson, 1998; Bordwell, et al., 1985; Cutting, 2005; Gibson, 1979; Lindgren, 1948).[2] In line with this ecological view of film cognition, explaining how a POV (point-of-view) shot is easily comprehended by viewers, Noël Carroll (1993) and Tim J. Smith (2012) argue that it mirrors natural attentional shifts between a looker and an object. Gaze following (looking where someone else is looking) emerges in infancy as early as 6 months of age to targets within a baby's own visual field (D'Entremont, Hains & Muir, 1997) and within the first year to targets more broadly (Butterworth & Jarrett, 1991; Corkum & Moore, 1998). By 12 months, infants will turn to see what another is looking at (Tomasello, Kruger, & Ratner, 1993). Adults, however, spontaneously monitor a person's eyes and use gaze direction to support inferences about his or her intentions, emotions, attention, knowledge states and likely future actions. Indeed, although other cues such as head orientation, body posture, or even pointing gestures may also provide important information in the determination of where gazers are directing their attention, the information from gaze cues have been shown to be exceptionally powerful (Perrett et al., 1992). The location of objects in the proximity of the viewer can also influence interpretation of gaze direction (Lobmaier et al., 2006) however in a Kuleshov-type sequence, these are unlikely to influence responses unless participants perceive spatio-temporal continuity between the adjacent (gazer, object) shots.

[2] Please see Smith, 2012, Smith, Levin and Cutting, 2012; Smith, 2012 and Ildirar & Schwan, 2014 for further discussions about perception of continuity in films.

(Figure 2 around here)

According to Noël Carroll (1993) the fact that a head movement is replaced with an edit does not matter because 'it is the endpoints of the activity, and not the space between, that command our attention' (Carroll, 1993, p. 128). Per Persson (2003) develops this theory by describing the POV structure as an instance of deictic gaze or joint visual attention. According to Persson (2003), in a POV scenario, the presentation of the object involves an unnatural 'jump' from one optical perspective/camera position to another. He suggests some conditions to increase the likelihood for the viewer to make a 'POV inference", and the first of these conditions is that the gazer should not look directly into the camera (cf. the so-called 'fourth wall' rule)[3]. Perhaps crucially, the original Kuleshov sequences do not follow such suggestion. Moreover, since the present study aims to replicate the original sequences as closely as possible, in our core stimuli the gazer will look directly into the camera.

The technique of direct address—when a character looks to the audience - is rare in fictional cinematic discourse except in instances of comedy (Renov, 2004, p.30). However, this technique has become increasingly popular with documentary film makers since the 1990s. It is believed to stand-in for eye-contact in daily life and increase the sense of intimacy as well as confrontation (Rosenheim 1996, p.221). Interestingly, a study investigating perceptions of credibility during testimony reported that witnesses who averted their gaze were perceived to be less credible and were more likely to be associated with a guilty verdict (Hemsley & Doob, 1978). Others have since also found that maintaining eye contact with an interviewer facilitates deception detection (Vrij, Mann, Leal, & Fisher, 2010). It follows then,

[3] Please see Barratt et. al., 2016 for the summary of the Persson's theory.

that looking directly into the camera might have an effect (positive or negative) on perception of continuity and emotion: both components of the Kuleshov effect.

Kuleshov-Type Sequence as a Place for Emotion Seeds to Sprout

In everyday life, face stimuli are rarely perceived in isolation and the context in which they appear can be very informative. Researchers have explored three types of context effects on facial emotion perception: (a) stimulus based context, in which a face is physically presented with other sensory input that has informational value; (b) perceiver based context, in which processes within the brain or body of a perceiver can shape emotion perception; and (c) cultural contexts that affect either the encoding or the understanding of facial actions (Barrett, Batja , and Gendron, 2011). The Kuleshov experiment deals with stimulus-based context.

Emotion perception studies investigating the influence of stimulus-based context have shown that facial expression judgments are influenced by any number of cues, including descriptions of the social situation (e.g., Carroll & Russell, 1996), voices, body postures, and visual scenes (e.g., Aviezer et al., 2008; Righart & de Gelder, 2008; for reviews, see Barrett, Mesquita & Gendron, 2011 and de Gelder et al., 2006), and even other faces (e.g., Masuda et al., 2008). For example, scowling faces (posed, exaggerated facial expressions of anger) are more likely to be perceived as fearful when paired with a description of danger (Carroll & Russell, 1996, Study 1) or disgusted when paired with a body posture involving a soiled object (Aviezer et al., 2008, Study 1). Aviezer and colleagues (2008) propose a model of context effects using the metaphor of ‘‘emotion seeds.’’ They suggest that the same perceptual information might be shared by different facial expressions (i.e., emotion seeds) and lie dormant in isolated faces, but be activated by appropriate context. If a given context activates a facial expression that shares enough emotion seeds with the expression displayed

by a target face, these seeds will ‘‘sprout’’ and override the original expression of the target face. By contrast, an equally powerful context will have little impact if its associated facial expression shares few emotion seeds with the expression of the target face (Aviezer, 2008, p.9). In case of naïve viewers viewing a Kuleshov sequence, we speculate that the sprout of seeds might function to not only help them perceive an expression on an otherwise expressionless face, it could also help them link the discontinuous shots.

Method

Participants

Forty participants (half female, 56–72 years old, M = 64.1 years) took part in the study. All subjects gave informed consent and the study was approved by the Research Ethics Committee of the University Hospital of Istanbul University. The experimental group (20 participants, half female, 58–72 years, M= 66.4 years) knew of the existence of television and had some abstract ideas about it, but had no prior direct experience with the medium. This group lived in small isolated houses in the mountains south of Isparta, Turkey that had only recently been connected to the electrical grid. All these of participants had some photos (mostly head shots of their children or grandchildren) and four had radios with a very limited broadcast range. Many assumed that television is a “visual radio” with programs that showed pictures of the people who speak or sing on the radio. Seven of the group was illiterate and the average years of schooling was 1.95 years.

The control group (half female, 56–72 years, M = 61.9 years) were from a similar geographic and cultural background as the experimental group. Critically, these participants all had some experience with television. They spoke the same dialect and had a similar lifestyle as the experimental group, but with a little more access to luxuries. Three of them were illiterate and the average education level was 3.1 years. This control group was

significantly younger than the experimental group, $F(57,2) = 3.7$, $p = .03$, but there was no significant difference in educational level, $x^2(4) = 4.48$, $p = .3$.

Stimuli

Two sets of video clips were produced, which each contained six two-shot sequences of eight second durations (see Table 1). In Set A, each sequence started with an expressionless man's face, which was followed by an image of a plate of soup, a gravestone, and a little girl. In Set B, the sequences matched the structure of those in Set A, but here the facial expression of each man matched the intercut images, i.e., he licked his lips and gulped to express hunger (when preceding the soup image), looked sad (when preceding the gravestone) and smiled (when preceding the little girl). Two versions of each set were created featuring different actors. We independently validated perception of these expressions (i.e., as showing someone neutral, hungry, sad, happy) with a large separate group of undergraduate students (n=80). To replicate the conditions in Kuleshov's original experiment, in both clips the actors looked directly into the camera, the sequences were grayscale and there was no sound.

An additional sequence was produced during testing in the field following responses from the first three experimental (naïve) participants that strongly signaled they were not making any connections between the intercut images. In light of this, we made an alternate version of the hunger sequence in Set A where the actor was replaced with a shot of an old woman looking down and a plate of soup on a floor table, which is where these participants tend to eat their own meals.

(Table 1 around here)

Procedure

Participants were tested individually in their homes in sessions lasting 30 - 60 minutes. In order to check for possible auditory, visual, or cognitive deficits, participants were asked to describe their present situation, i.e., what they could see outside the window. They were also interviewed about their experience with and their knowledge about television and films. No participants were excluded on the basis of these discussions.

After these questions, a laptop with a 17.3-inch display was presented to participants (viewing distance ~ 60 cm). Participants were told that they would see something on the display and be asked to describe it, as they had previously described their present (real-life) situation. The video sequences were shown in a fixed order (as in Table 1) with a short break after each presentation to answer questions from the experimenter. The first question was always "Could you please tell me what you have seen?" If their answer clearly indicated an understanding of spatiotemporal continuity and/or the Kuleshov effect (e.g., "I saw a man smiling at the baby across him") no further questions were asked regarding spatio-temporal continuity perception. When the participants mentioned just one of the shots (e.g., "I saw a man looking at me"), they were always asked what else they saw, which usually led them to talk about the content of the other shot (e.g., "There was a man first. Then he disappeared and there appeared a stewpan"). If the answer did not mention any connection between the shots (e.g., "I saw a gravestone too"), follow-up questions were also asked (e.g. "Where was the gravestone?") until their perception of the edited sequence was clear. All the participants were also asked how the person on the screen was feeling.

Coding and Analysis

All sessions were video recorded, transcribed and then double coded (reliability, Cohen's kappa coefficient > .92) using the qualitative analysis program Atlas-ti. Each participant's qualitative responses to each clip was numerically classified. When there was no spatio-temporal linkage between the camera shots (i.e., no sense that the person in the first

shot was in the same place or time as the objects in the second shot) they received a score of 0. When they *did* make a clear spatio-temporal link between the shots they received a score of 1. When participants demonstrated a clear Kuleshov effect (i.e., perceived variation in the (neutral) facial expression of the first shot when it interacted with the content in the second shot) they received a score of 2. After the coding process, the data were transferred from Atlas-ti to SPSS and differences in the frequencies between the first-time viewers and the experienced viewers were tested for significance by Fisher's exact test.

Results

The percentage values reported below reflect participants responses averaged across the two identities that appeared in Sets A and B. Differences in frequencies between the first-time viewers and the experienced viewers were tested for significance using Fisher's exact test.

(Table 2 around here)

Set A (neutral faces)

First time viewers. The first-time viewers interpreted all the sequences in Set A as independent images. Responses in this group did not suggest any spatio-temporal linkage between the shots or the existence of a Kuleshov effect on the perceived expression of the neutral face. A typical response was that there was a man looking towards the viewer sitting in silence that came and went. When asked what else they saw, participants commented that the man disappeared and then something else appeared: a plate (often described as something bigger, e.g., a cooking pot or saucepan), a gravestone, or a little girl. When asked additional questions to probe their perception of these sequences, their responses revealed very limited consideration of context or interaction between the shots. With regards to the perceived spatial location of these objects (e.g when asked, "where was the plate/man?") they responded that the plate "should be in the kitchen" or "on the stove" or "how can I know it? It appeared

there (pointing the screen)". When asked what the man was feeling or thinking, first time viewers said that "they cannot know" that or "he was looking with empty eyes". When asked "was the little girl alone?" all participants answered "yes" adding that they did not see her parents next to her. The customized additional video clip added during testing, which featured a face with directed gaze (looking in the direction of the soup) helped the first-time viewers link the shots spatio-temporally. All of them reported that she was sitting at a floor table and waiting. The reasons provided for her waiting were diverse and related mostly to the individual backgrounds of the first-time viewers. For example, one female participant said that the woman in the video clip was afraid of her husband's anger since she did not know if he would like the meal. Given that these attributions regarding the woman's emotion were elicited in a perceiver-based rather than a stimulus-based context, this was not considered evidence of the Kuleshov effect.

Experienced viewers. In contrast to the first time viewers, 100% of the experienced viewers constructed spatio-temporal links between the shots in the Set A sequences. A Kuleshov effect was also observed for 55% of participants in the gravestone sequence.

For the soup sequence, 100% of participants reported that they saw a man with a meal in front of him with many (65%) also making a forward inference and saying that the man will eat the meal. When asked about how he was looking and feeling, 30% of participants said that he looked indecisive and was thinking whether he should eat the meal and 45% of them said that he was waiting for someone to start eating. The remaining 25% said "nothing special…he will just eat the meal". Here the absence of motion through the cuts led the viewers to seek an explanation for the two shots (i.e., the meal would be eaten by the actor). This expectation may be explained by the dramatic principle called 'Chekhov's gun'. Here, every element in a narrative is required to be irreplaceable (Bill, 1987). Thus just as whenever

you introduce a rifle in the first chapter it must go off in the second chapter (to give Chekhov's example), it seems that if you show a meal in the first shot of an edited sequence it must be eaten in the second shot.

For the gravestone sequence, 100% of experienced viewers made spatio-temporal links between the shots and 55% demonstrated a Kuleshov Effect. That is, they all said that the man was standing in front of a gravestone and when asked how he was feeling, 55% of them said that he looked sad/sorry. Other responses were that he was praying (15%) or keeping a minute of silence (20%), which might also be considered as interpretation of sadness since these are what people do in memory of people died. Only 10% of experienced viewers said that the person was feeling nothing.

For the child sequence, once again 100% of participants made spatio-temporal links between the face and the second image. All of the experienced viewers reported that they saw a man and a girl. When asked where they were, participants said that they must be at home or at school. No participants showed a clear Kuleshov effect. Forty-five percent of participants said that he felt "nothing", 20% said that he was miles away and thinking of something else. Interestingly, 25% of participants linked this sequence with the gravestone sequence (that preceded it) by saying that the man was trying to forget someone who had been lost by thinking that the life goes on.

For the old woman sequence, all participants reported that she was waiting before eating her meal. The reasons for waiting were several: allowing the meal to get cold (10%), expecting someone to come (20%) or that she just did not have appetite (45%). The rest did not make an explanation. When it was asked what she felt or thought, the most frequently made interpretation was "who knows what problem she has". Just as for the other 'soup' video clip (showing the man actor), however, no one inferred that she was hungry.

Set B (expressive faces)

First time viewers. Even with these emotion-congruent stimuli, first time viewers rarely constructed any links between the camera shots. It was only for the graveyard sequence that there was any evidence of any interaction. Critically, however, this did not constitute a full spatio-temporal association. Rather, participants said that they thought the man was sorry for his loss but did not seem to perceive him to be spatially located in the graveyard. When they were asked where he was, he was not reported to be across or next to gravestone but rather "here", looking at us. When they were further probed regarding where the gravestone was, participants responded that "it was gone". In the other sequences, even this limited interaction was not observed. For the soup sequence for example, participants described the man to be licking his lips/gulping (0% said he looked hungry) and then said that the plate (or pot/well/hole/pool) appeared *"again*". When asked the reason for this man's behavior, they said that they cannot know it. For the child sequence, the two shots were also interpreted as two independent pictures. The little girl and the man were said to be looking happy, but no participants commented that they were together.

Experienced viewers. Descriptions of the soup, graveyard, playing child and old woman (with directed gaze) sequences all indicated that 100% of experienced viewers made clear spatio-temporal associations between these shots. Furthermore, most of these participants perceived the emotions of the persons in the predicted manner: describing the man as hungry in the soup condition (95% for Actor A and 100% for Actor B), sad in the gravestone condition (100% for both actors) and happy in the child condition (100% for both actors).

(Table 3 around here)

Discussion

The "Kuleshov Effect" reflects a conscious connection and subsequent interaction between edited camera shots made by an observer. The camera shots used in the sequences typically associated with this effect are not connected to each other with commonalities on a perceptual level but rather through intentions, motivations and emotions. In other words, any continuity between juxtaposed shots is an illusion created in the mind of a viewer and the landscape in which both shots are located in is an artificial one: existing outside of reality. The present study investigated whether first-time viewers construct spatio-temporal relations between the shots like experienced viewers, i.e., forging a narrative connection between them and conceiving of the artificial landscape created in the video clips. Here, we coded naïve and experience viewers' responses to classic Kuleshov experiment sequences in order to establish whether or not there are differences in how first-time film viewers spontaneously connect edited shots and generate the Kuleshov Effect.

The current study did not address different theories of emotion; the existence, discreteness of specific emotions or other related topics[4]. A person's ability to make sense of facial expressions is affected by several factors, which we attempted to control as much as possible with in the present study. Responses from the experienced viewer participant group validated our chosen stimulus set. These participants all connected the shots on spatio-temporal level and had no problem identifying the facial expressions used in the Set B sequences. Moreover, even the first time viewers accurately categorized the emotions in the

[4] For cutting edge theories of emotion, see (Moors et al. 2013) and for current debates on universality of emotion recognition see (Nelson and Russell 2013).

446 happiness and sadness conditions but not the hunger condition, which could reflect this latter
447 judgment not being one of the 'basic' or universal emotions.

448 Our results reveal that first-time viewers do not demonstrate either of the two key
449 components of the Kuleshov effect. Despite an intact ability to perceive and understand the
450 content of each shot, they perceived them to be wholly separate from each other and did not
451 relate them spatio-temporally. Even when the coherent facial expressions were juxtaposed
452 with the causes of such expressions, they still considered them as if they were independent
453 photographs: a visual format they are familiar with. First-time viewers do not seem to have
454 the notion of what constitutes a film, i.e., sequences of shots that are linked in coherent ways.
455 In the sadness condition, for example, they said that the person is sad because of someone he
456 had lost (in relation to the gravestone shot), but crucially there was no indication that they
457 thought that the sad person and the gravestone were in the same place at the same time. The
458 image of the person was not "here" any more as the image of gravestone. These results are
459 consistent with the results of a study for young children when looking at picture books
460 (Berman 1988) that suggest that once a page is turned, a new story begins. Berman (1988)
461 concluded that the narrative abilities that function to allow children to link such events are
462 constrained by broader development in cognition, expressive language abilities, and their
463 familiarity with narrative norms in a literate society.

464 Somewhat surprisingly, the customized additional video clip added during testing
465 revealed that first-time viewers *can* connect edited sequences spatio-temporally under at least
466 some conditions. For example, when a person's gaze in the first shot is coherent with the
467 location of the depicted object in the second shot. Here, the eye-line match filmic equivalents
468 of joint attention (acquired in early childhood, e.g., (Moore and Dunham 2014) may have
469 provided an instance of a conceptual relation that was clear enough for even naïve viewers to

interpret. Eye-line matches, in other words, appeared to open the eyes of first-time viewers to the artificial landscape created in the video clip. Unfortunately, there was no scope for interpretation of the facial expression of the lady depicted because her face was not readable (head and eyes were turned downwards) preventing evaluation of the second component of the Kuleshov effect.

The 'classic' Kuleshov effect was clearly observed for experienced viewers only in sadness condition. Here, participants reported that the man standing in front of the gravestone was *sad* for his loss, although the footage showed the same neutral expression face that was juxtaposed with the shots of soup and the little girl. It could be argued that the image of the gravestone is much intense and salient than images of a plate of soup or a cute child playing. This study, however, followed the procedures described in other studies of the Kuleshov effect, for comparability with this previous work. It is possible that these participants' interpretations of the emotional state of the faces shown before the plate of soup could also be considered evidence of the Kuleshov effect in action. Although there was no clear attribution of a specific emotional or mental state, the experienced viewers tried to find an explanation for what caused the man to not eat the soup in front of him. Thirty percent of them said that the man was unsure whether or not he should eat it, and forty-five percent thought he was waiting for someone else.

When considering participants' responses to the video sequences with the little girl, it may be helpful to consider that viewing one facial expression can shift the wider scale of judgment. That is, a strongly salient 'anchor' face can skew the emotion perceived in subsequent faces in the opposite affective direction (Russell and Fehr 1987), making a neutral face appear sad when presented after a happy face, or happy when presented after a sad face.

493 Thus, the happy face of the little girl in the test sequences might have biased participants
494 interpret the actor's facial expression to be relatively less happy, or even sad.

495 Prince and Hensley (1992) cited the naiveté of the early audiences as a possible reason
496 for discrepancies in the appearance of the Kuleshov effect with contemporary audiences. Our
497 results challenge this notion. They indicate that first time film viewers do not even link
498 intercut camera shots edited in sequence, let alone demonstrate the Kuleshov effect. We
499 propose, instead, that it is experienced viewers that are more likely to 'collaborate' with the
500 film-maker. That is, to try to understand their intentions and make sense of what they see
501 because they know that films are comprised of shots that come together to convey a narrative.
502 Such viewers contrast with naïve viewers who seem unaware of the existence of a film-maker
503 or a camera. It should be noted here that the experienced viewers in the present study (like the
504 first-time viewers) had no prior experience of taking part in research. Both participant groups
505 were first-time participants in a study and had no idea about what a *study* is. Even though the
506 experiment was explained, they supposed that they would simply watch videos without
507 realizing that they were purpose-made for research purposes.

508 It seems also worth mentioning here that the first-time viewers (mis)interpreted the
509 objects shown in close-up shots as things bigger than they really were (e.g. plate, hole) and
510 the people as sitting (only upper bodies were shown, in medium shots). These were also
511 evidence that the first-time viewers recruited for this particularly study had only a very basic
512 understanding of what film is. It was also interesting that neither first-time nor experienced
513 viewers made any comment on the black and white quality of the video clips. Further
514 research is needed to determine the role of such prior knowledge by explaining the concept of
515 film to first-time viewers. Further research is also needed to test the Kuleshov effect with
516 other images, e.g., those as perceptually salient as a gravestone, which might elicit stronger

emotions and modulate perception more powerfully. Direction of gaze and the order of the shots have also been identified as key variables that should also be taken into account in such work.

Filmography

Hemmings, D. 1978, Just A Gigolo, USA

Hitchcock, A. 1954, Rear Window, USA

Kuleshov, L. 1918, The Project of Engineer Prite, Rusia

Bibliography

Anderson, J.D., 1998. The reality of illusion: *An ecological approach to cognitive film theory*. SIU Press.

Aviezer, H., Hassin, R.R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M. and Bentin, S., 2008. Angry, disgusted, or afraid? *Studies on the malleability of emotion* perception. *Psychological Science*, 19(7), pp.724-732.

Bakeman, R. and Adamson, L.B., 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child development*, pp.1278-1289.

Barrett, L.F., Mesquita, B. and Gendron, M., 2011. Context in emotion perception. *Current Directions in Psychological Science*, 20(5), pp.286-290.

Barratt, D., Rédei, A.C., Innes-Ker, Å. and van de Weijer, J., 2016. Does the Kuleshov effect really exist? Revisiting a classic film experiment on facial expressions and emotional contexts. *Perception*, p.0301006616638595.

Berman, R.A., 1988. On the ability to relate events in narrative. *Discourse Processes*, 11(4), pp.469-497.

Bill, V.T., 1987. *Chekhov: The Silent Voice of Freedom*. Allied Books Limited.

Bordwell, D., Staiger, J., and Thompson, K. 1985. *The Classical Hollywood Cinema: Flm Style and Mode of Production to 1960*. London, UK: Routledge.

Bordwell, D., Thompson, K. and Ashton, J., 1997. *Film art: An introduction* (Vol. 7). New York: McGraw-Hill.

Butterworth, G. and Jarrett, N., 1991. What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British journal of developmental psychology*, 9(1), pp.55-72.

Carroll, J M, and J a Russell. 1996. "Do Facial Expressions Signal Specific Emotions? Judging Emotion from the Face in Context." *Journal of Personality and Social Psychology* 70 (2): pp.205–18.

Carroll, N. 1993. "Toward a theory of point-of-view editing: Communication, emotion, and the movies." *Poetics Today* pp.123-141.

Corkum, V. and Moore, C., 1998. The origins of joint visual attention in infants. *Developmental psychology*, 34(1), p.28.

Croy, H. 2009. *How Motion Pictures Are Made*. General Books LLC.

556 Cutting, J. E. 2005. "Perceiving scenes in film and in the world". Pp. 9-27 in *Moving Image*
557 *Theory: Ecological considerations*, eds. J. D. Anderson and B. F. Anderson. Carbondale,
558 IL: Southern Illinois University Press.

559 D'Entremont, B., Hains, S.M.J. and Muir, D.W., 1997. A demonstration of gaze following in
560 3-to 6-month-olds. I*nfant Behavior and Development*, 20(4), pp.569-572.

561 Elsaesser, T. 1990. *Early cinema: Space frame narrative*. BFI Publishing.

562 Fernandez-Dols, J.M., Sierra, B. and Ruiz- Belda, M., 1993. On the clarity of expressive and
563 contextual information in the recognition of emotions: A methodological
564 critique. *European Journal of Social Psychology*, 23(2), pp.195-202.

565 Forsdale, J.R. and Forsdale, L., 1970. Film literacy. *Educational Technology Research and*
566 *Development*, 18(3), pp.263-276.

567 de Gelder, B., Meeren, H.K., Righart, R., Van den Stock, J., van de Riet, W.A. and Tamietto,
568 M., 2006. Beyond the face: exploring rapid influences of context on face
569 processing. *Progress in brain research*, 155, pp.37-48.

570 Gibson, J.J., 2014. *The ecological approach to visual perception: classic edition.* Psychology
571 Press.

572 Goldberg, H.D., 1951. The role of" cutting" in the perception of the motion picture. *Journal*
573 *of Applied Psychology*, 35(1), p.70.

574 Goodenough, F.L. and Tinker, M.A., 1931. The relative potency of facial expression and
575 verbal description of stimulus in the judgment of emotion. *Journal of Comparative*
576 *Psychology*, 12(4), p.365.

Hemsley, G.D. and Doob, A.N., 1978. The Effect of Looking Behavior on Perceptions of a Communicator's Credibility1. *Journal of Applied Social Psychology*, 8(2), pp.136-142.

Hobbs, R., Frost, R., Davis, A. and Stauffer, J., 1988. How First-Time Viewers Comprehend Editing Conventions. J*ournal of Communication*, 38(4), pp.50-60.

Hochberg, J. and Brooks, V. 1996. "The Perception of Motion Pictures." *Cognitive Ecology*, pp.205–92.

Holland, Norman N. 1989. "Film Response From Eye To I-The Kuleshov Experiment." *South Atlantic Quarterly* 88 (2). Duke Univ Press Box 90660, Durham, NC 27708-0660 pp. 415–42.

Ildirar, S., Levin, D., Schwann, S., and Smith. T. 2010. " The Role Of Audio On The Perception Of Continuity By First-Time Viewers". Paper presented at the annual meeting for the The Society for Cognitive Studies of the Moving Image, Roanoke, Virginia, June 2-5

Ildirar, S. and Schwan, S., 2015. First- time viewers' comprehension of films: Bridging shot transitions. *British Journal of Psychology*, 106(1), pp.133-151.

Isenhour, J.P., 1975. The effects of context and order in film editing. *Educational Technology Research and Development*, 23(1), pp.69-80.

Kuleshov, Lev Vladimirovich. 1974. *Kuleshov on Film: Writings*. Univ of California Press.

Levaco, R., 1974. Eikhenbaum, inner speech and film stylistics. *Screen*, 15(4), pp.47-58.

596 Levin, D.T. and Simons, D.J., 1997. Failure to detect changes to attended objects in motion
597 pictures. *Psychonomic Bulletin & Review*, 4(4), pp.501-506.

598 Lindgren, E. 1948. *The Art of the Film*. London, UK: George Allen and Unwin Limited.

599 Lobmaier, J.S., Fischer, M.H. and Schwaninger, A., 2006. Objects capture perceived gaze
600 direction. *Experimental Psychology*, 53(2), pp.117-122.

601 Mangun, G. R. 2008. "Looking Inward: The Mind's Eye Focuses on Mental Representations."
602 *Frontiers in Neuroscience* 2 (2). Frontiers Research Foundation: p. 133.

603 Masuda, T., Ellsworth, P.C., Mesquita, B., Leu, J., Tanida, S. and Van de Veerdonk, E., 2008.
604 Placing the face in context: cultural differences in the perception of facial
605 emotion. *Journal of personality and social psychology*, 94(3), p.365.

606 Messaris, P. 1994. *Visual" literacy": Image, mind, and reality*. Westview Press.

607 Mobbs, D., Weiskopf, N., Lau, H.C., Featherstone, E., Dolan, R.J. and Frith, C.D., 2006. The
608 Kuleshov Effect: the influence of contextual framing on emotional attributions. *Social*
609 *cognitive and affective neuroscience*, 1(2), pp.95-106.

610 Moore, C. and Dunham,P. 2014. *Joint Attention: Its Origins and Role in Development*.
611 Psychology Press.

612 Moors, A., Ellsworth, P.C., Scherer, K.R. and Frijda, N.H., 2013. Appraisal theories of
613 emotion: State of the art and future development. *Emotion Review*, 5(2), pp.119-124.

614 Munsterberg, H. 1970. *The photoplay: A psychology study.*

615 Nakamura, M., Buck, R. and Kenny, D.A., 1990. Relative contributions of expressive
616 behavior and contextual information to the judgment of the emotional state of
617 another. *Journal of personality and social psychology*, 59(5), p.1032.

618 Nelson, N.L. and Russell, J.A., 2013. Universality revisited. *Emotion Review*, 5(1), pp.8-15.

619 Ondaatje, M., 2004. *The conversations: Walter Murch and the art of film editing.*

620 Pearson, R. and Simpson, P. eds., 2005. *Critical dictionary of film and television theory*.
621 Routledge.

622 Perrett, D.I., Hietanen, J.K., Oram, M.W., Benson, P.J. and Rolls, E.T., 1992. Organization
623 and functions of cells responsive to faces in the temporal cortex [and
624 discussion]. Philosophical Transactions of the Royal Society of London B: *Biological*
625 *Sciences*, 335(1273), pp.23-30.

626 Persson, P., 2003. *Understanding cinema: A psychological theory of moving imagery*.
627 Cambridge University Press.

628 Prince, S. and Hensley, W.E., 1992. The Kuleshov effect: Recreating the classic
629 experiment. *Cinema Journal*, 31(2), pp.59-75.

630 Pudovkin, V.I., 2013. *Film Technique and Film Acting-The Cinema Writings of VI Pudovkin*.
631 Read Books Ltd.

632 Renov, M., 2004. *The subject of documentary* (Vol. 16). University of Minnesota Press.

633 Righart, R. and De Gelder, B., 2008. Recognition of facial expressions is influenced by
634 emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience*, 8(3), pp.264-272.

635 Rosenheim, S., 1996. Interrotroning history: Errol Morris and the documentary of the
636 future. *The Persistence of History: Cinema, Television, and the Modern Event*, pp.219-
637 234.

638 Russell, J.A. and Fehr, B., 1987. Relativity in the perception of emotion in facial
639 expressions. J*ournal of Experimental Psychology: General*, 116(3), p.223.

640 Sharff, S. 1997. The Art of Looking in Hitchcock's R*ear Window*. Hal Leonard Corporation.

641 Schwan, S. and Ildirar, S., 2010. Watching Film for the First Time How Adult Viewers
642 Interpret Perceptual Discontinuities in Film. *Psychological Science*.

643 Sklar, R. 2002. *A World History of Film*. Prentice Hall.

644 Smith, T. J. 2012. "The attentional theory of cinematic continuity". *Projections*, 6(1), pp.1-27.

645 Smith, T.J., Levin, D. and Cutting, J.E., 2012. A window on reality perceiving edited moving
646 images. *Current Directions in Psychological Science*, 21(2), pp.107-113.

647 Tomasello, M., Kruger, A.C. and Ratner, H.H., 1993. Cultural learning. B*ehavioral and brain*
648 *sciences*, 16(03), pp.495-511.

649 Truffaut, F. 1984. *Hitchcock* (rev. ed.). In collaboration with H. G. Scott. New York, NY:
650 Simon & Schuster, Inc.

651 Vrij, A., Mann, S., Leal, S., & Fisher, R. 2010. 'Look into my eyes': can an instruction to
652 maintain eye contact facilitate lie detection?. *Psychology, Crime & Law*, 16(4), pp. 327-
653 348.

Wallbott, H. G. 1990. *Mimik Im Kontext: Die Bedeutung Verschiedener Informationskomponenten Für Das Erkennen von Emotionen*. Verlag für Psychologie

Walden, T. A., and Tamra A. Ogan. "The development of social referencing." *Child development* (1988): pp.1230-1240.

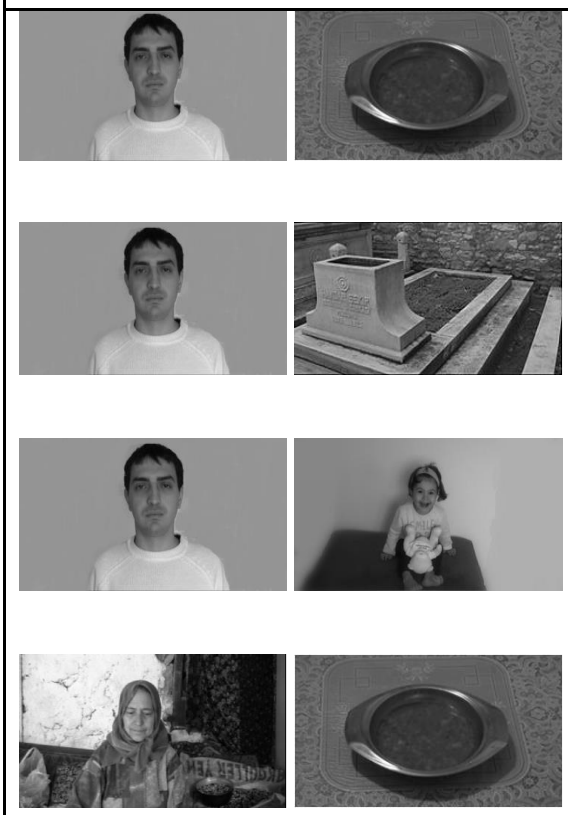
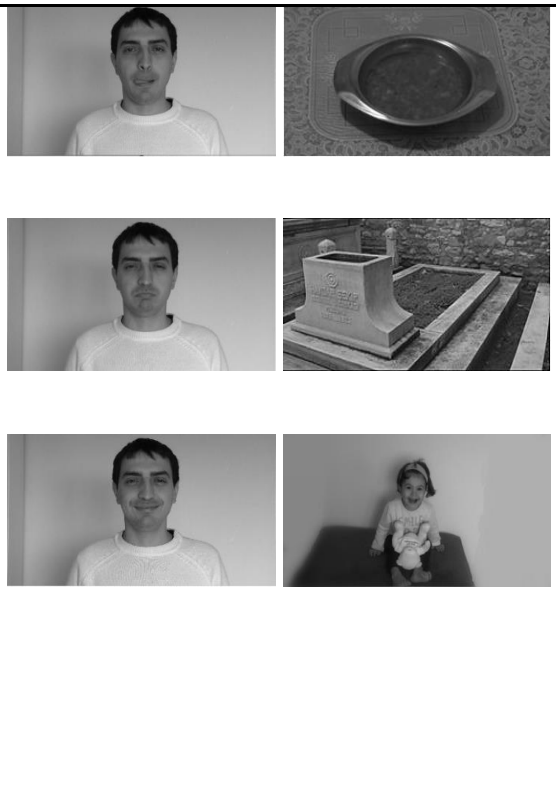
Set A (neutral facial expression)	Set B (consistent facial expression)
	

Table 1. An overview of the two-shot sequences produced for this study. Each eight second sequence in Set A featured a face with a neutral facial expression and those in Set B featured an overt expression that matched the following image. Alternate versions of each sequence were created with a second actor. Set B included an additional sequence added in the field that was intended to more closely mirror the conditions in participants' lives (old woman looking down).

660

Film Sequence	**Spatiotemporal continuity perception (%)**			**Kuleshov Effect perception (%)**		
	Viewer			**Viewer**		
	Naïve	Experienced	Group comparison[A]	Naïve	Experienced	Group comparison[A]
Man + Soup	0	100	p<.001	0 0	0(hungry) 30(indecisive)	p=1 p= < .01
Man + Gravestone	0	100	p<.001	0	55	p<.05
Man + Baby	0	100	p<.001	0	0	p=1
Local Lady (looking down) + Soup	100	100	p=1	0	0	p=1

A. Fisher's exact test

Table 2: Perception of spatiotemporal continuity and the Kuleshov Effect across groups for the first set of film sequences: intercut faces with neutral expressions.

Film Sequence	**Spatiotemporal continuity perception (%)**			**Correct interpretation of depicted emotion (%)**		
	Viewers			Viewers		
	Naïve	Experienced	Group Difference[A]	Naïve	Experienced	Group Difference[A]
Hungry man + soup	0	100	p<.001	0	95	p<.001
Sad man + gravestone	0	100	p<.001	100	100	p=1
Happy man + baby	0	100	p<.001	100	100	p=1

A. Fisher's exact test

Table 3: Perception of spatiotemporal continuity and the Kuleshov Effect across groups for the second set of film sequences: with edit-congruent facial expressions.

Figure 1. Example of sequence from Ildirar and Schwan (2014). First-time viewers could not construct a spatiotemporal relationship between adjacent shots.

Figure 2. Example of sequence from Ildirar and Schwan (2014). Salient gaze cue helped first-time viewers to construct a spatiotemporal relationship between adjacent shots.

673

674